The Sample Complexity of Randomized Methods for Analysis and Design of Uncertain Systems

T. Alamo [a], R. Tempo [b], A. Luque [a], D.R. Ramirez [a]

[a] *Departamento de Ingeniería de Sistemas y Automática, Universidad de Sevilla, Escuela Superior de Ingenieros, Camino de los Descubrimientos s/n, 41092 Sevilla. Spain*

[b] *CNR-IEIIT, Politecnico di Torino, Corso Duca degli Abruzzi 24, Torino 10129, Italy*

Abstract

In this paper, we study the sample complexity of probabilistic methods for uncertain systems. In particular, we show the role of the binomial distribution for some problems involving analysis and design of robust controllers with finite families. We also address the particular case in which the design problem can be formulated as an uncertain convex optimization problem. The second main contribution of the paper is to study a general class of sequential algorithms which satisfy the required specifications using probabilistic validation methods and, at each iteration of the sequential algorithm, a candidate solution is probabilistically validated. The results of the paper provide the sample complexity which guarantees that the obtained solutions meet some pre-specified probabilistic specifications.

Key words: randomized algorithms, probabilistic robustness, uncertain systems, sample complexity

1 Introduction

Research on probabilistic analysis and design for systems and control has significantly evolved. In recent years, key areas where we have seen convincing developments include uncertain and hybrid systems [31]. A key feature of this approach is the use of the theory of rare events and large deviation inequalities which suitably bound the tail of the probability distribution. These inequalities are crucial in the area of Statistical Learning Theory [32], which has been utilized for feedback design of uncertain systems [33].

Design in the presence of uncertainty is of major relevance in different areas. Unfortunately, the related semi-infinite optimization problems are often NP-hard problems, and this may seriously limit their solution from the computational point of view [7]. There are two approaches to resolve this NP-hard issue. The first approach is based on the computation of deterministic relaxations of the original problem which are normally solved in polynomial time. However, this might lead to overly conservative solutions [27]. An alternative paradigm is to assume that the plant uncertainty is probabilistically described and a randomized algorithm may be developed to compute, in polynomial time, a solution with some given properties stated in terms of the probability of error [31], [33].

A survey on probabilistic methods can be found in [13]. Two complementary approaches, non-sequential and sequential, have been proposed. A classical approach for non-sequential methods is based upon Statistical Learning Theory [32], [33]. Subsequent work along this direction includes [21], [34], [35], [2]. In [4], [3] and [23] the case in which the design parameter set has finite cardinality is analyzed. The advantage of these methods is that the problem under attention may be non-convex. For convex optimization problems, a successful non-sequential paradigm, denoted as the scenario approach, has been introduced in [9] and [10], see also [14], [15], [12] and [4] for related results.

For non-sequential methods the original robustness problem is reformulated as a single optimization problem with sampled constraints which are randomly generated. A relevant feature of these methods is that they do not require any validation step and the sample complexity is defined a priori. The main result of this line

Email addresses: `alamo@cartuja.us.es` (T. Alamo), `roberto.tempo@polito.it` (R. Tempo), `amalia@cartuja.us.es` (A. Luque), `danirr@cartuja.us.es` (D.R. Ramirez).

of research is to derive explicit lower bounds to this required sample size. However, the obtained explicit sample bounds can be overly conservative because they rely on a worst-case analysis and grow (at least linearly) with the number of decision variables.

For sequential methods, the resulting iterative algorithms are based on stochastic gradient [8], [26], ellipsoid iterations [20], [24], [29], or analytic center cutting plane methods [11], [28], see also [1] for other classes of sequential algorithms. Convergence properties in finite-time are in fact one of the focal points of these papers. Various control problems have been solved using these sequential randomized algorithms, including robust LQ regulators [26], switched systems [22], and uncertain linear matrix inequalities (LMIs) [8]. Sequential methods are mostly used for uncertain convex feasibility problems because the computational effort at each iteration is affordable. However, they have been studied also for non-convex problems, see [2], [19].

The common feature of all these sequential algorithms is the use of the validation strategy presented in [24]. The candidate solutions provided at each iteration of these algorithms are validated using a validation set which is drawn according to the probability measure defined in the uncertain set. If the candidate solution satisfies the design specifications for every sampled element of this validation set, then it is classified as probabilistic solution and the algorithm terminates. The main point in this validation scheme is that the cardinality of the validation set increases very mildly at each iteration of the algorithm. The strategy guarantees that, if a probabilistic solution is obtained, then it meets some probabilistic specifications. A similar approach, introduced in [17], has been presented in [13] in the context of sequential algorithms. The contribution is a reduction on the cardinality required for the validation sets. See also [5] for related results.

In this paper, we derive the sample complexity for various analysis and design problems related to uncertain systems. In particular we provide new results which guarantee that the tail of the binomial distribution is bounded by a pre-specified value. These results are then applied to the analysis of worst-case performance and constraint violation. With regard to design problems we consider the case of finite families and the special case when the design problem can be recast as a robust convex optimization problem.

The second main contribution of this paper is to propose a relaxed validation scheme which allows the candidate solution to violate the design specifications for one (or more) of the members of the validation set. The idea of allowing some violations of the constraints is not new and can be found, for example, in the context of system identification [6], chance-constrained optimization [15] and Statistical Learning Theory [2]. This scheme makes sense in the presence of soft constraints or when a solution satisfying the specifications for all the admissible uncertainty realizations cannot be found.

The rest of the paper is organized as follows. In the next section, we first introduce the problem formulation. In Section 3, we provide bounds for the binomial distribution which are used in Section 4 to analyze the probabilistic properties of different schemes involving randomization. In Section 5, we introduce the proposed family of probabilistically validated algorithms. The sample complexity of the validating sets is analyzed in Section 6. A comparison with the validation scheme presented in [24] is provided in Section 7. Section 8 studies how to use the results of the paper in the context of non-sequential randomized algorithms. A numerical example in which the different schemes are used to address a robust identification problem is presented in Section 9. The paper ends with a section of conclusions.

2 Problem Statement

We assume that a probability measure $\Pr_{\mathcal{W}}$ over the sample space $\mathcal{W}$ is given. Given $\mathcal{W}$, a collection of N independent identically distributed (i.i.d.) samples $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ drawn from $\mathcal{W}$ belongs to the Cartesian product $\mathcal{W}^N = \mathcal{W} \times \cdots \times \mathcal{W}$ (N times). Moreover, if the collection w of N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ is generated from $\mathcal{W}$ according to the probability measure $\Pr_{\mathcal{W}}$, then the *multisample* w is drawn according to the probability measure $\Pr_{\mathcal{W}^N}$. The scalars $\eta \in (0,1)$ and $\delta \in (0,1)$ denote probabilistic parameters called accuracy and confidence, respectively. Furthermore, $\ln(\cdot)$ is the natural logarithm and e is the Euler number. For $x \in \mathbb{R}$, $x \geq 0$, $\lfloor x \rfloor$ denotes the largest integer smaller than or equal to x; $\lceil x \rceil$ denotes the smallest integer greater or equal than x. For $\alpha \in \mathbb{R}$, $\alpha > 1$, $\xi(\alpha)$ denotes the Riemann zeta function (i.e. $\xi(\alpha) = \sum\limits_{k=1}^{\infty} \frac{1}{k^{\alpha}}$).

In a robustness problem, the design parameters and auxiliary variables are parameterized by means of a decision variable vector θ, which is denoted as "design parameter", and is restricted to a set Θ. Furthermore, the uncertainty w is bounded in the set $\mathcal{W}$ and represents one of the admissible uncertainty realizations. We also consider a binary measurable function $g : \Theta \times \mathcal{W} \to \{0,1\}$ and a real measurable function $f : \Theta \times \mathcal{W} \to \mathbb{R}$ which helps to formulate the specific design problem under attention. More precisely, the binary function $g : \Theta \times \mathcal{W} \to \{0,1\}$, is defined as

$$g(\theta, w) := \begin{cases} 0 & \text{if } \theta \text{ meets design specifications for } w \\ 1 & \text{otherwise,} \end{cases}$$

where design specifications are, for example, H_∞ norm

bounds on the sensitivity function, see specific examples in [31], or approximation of functions, see the numerical example in Section 9.

Given $\theta \in \Theta$, the constraint $g(\theta, w) = 0$ is satisfied for a subset of $\mathcal{W}$. This concept is rigorously formalized by means of the notion of "probability of violation", which is now introduced.

Definition 1 *[probability of violation] Consider a probability measure* $\Pr_{\mathcal{W}}$ *over* $\mathcal{W}$ *and let* $\theta \in \Theta$ *be given. The probability of violation of* θ *for the function* $g : \Theta \times \mathcal{W} \to \{0, 1\}$ *is defined as*

$$E(\theta) := \Pr_{\mathcal{W}}\{\, w \in \mathcal{W} \;:\; g(\theta, w) = 1 \,\}.$$

Using this notion we study the robust optimization problem

$$\min_{\theta \in \Theta} J(\theta) \;\text{ subject to }\; E(\theta) \le \eta \tag{1}$$

where $J : \Theta \to (-\infty, \infty)$ is a measurable function which normally represents the controller performance. Given accuracy $\eta \in (0, 1)$ and confidence $\delta \in (0, 1)$, the main point of the probabilistic approach is to design an algorithm such that any probabilistic solution $\hat{\theta}$ obtained running the algorithm satisfies $E(\hat{\theta}) \le \eta$ with probability no smaller than $1 - \delta$.

Even in analysis problems when $\theta \in \Theta$ is given, it is often very hard to compute the exact value of the probability of violation $E(\theta)$ because this requires to solve a multiple integral with a non-convex domain of integration. However, we can approximate its value using the concept of empirical mean. For given $\theta \in \Theta$, and multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ drawn according to the probability measure $\Pr_{\mathcal{W}^N}$, the empirical mean of $g(\theta, w)$ with respect to w is defined as

$$\hat{E}(\theta, \mathrm{w}) := \frac{1}{N} \sum_{i=1}^{N} g(\theta, w^{(i)}).$$

Clearly, the empirical mean $\hat{E}(\theta, \mathrm{w})$ is a random variable. Since $g(\cdot, \cdot)$ is a binary function, $\hat{E}(\theta, \mathrm{w})$ is always within the closed interval $[0, 1]$.

The power of randomized algorithms stems from the fact that they can approximately solve non-convex design problems (with no violation) of the type

$$\min_{\theta \in \Theta} J(\theta) \;\text{ subject to } g(\theta, w) = 0, \text{ for all } w \in \mathcal{W}. \tag{2}$$

In this setting, we draw N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$ and solve the sampled optimization problem

$$\min_{\theta \in \Theta} J(\theta) \;\text{ subject to } g(\theta, w^{(\ell)}) = 0,\; \ell = 1, \ldots, N. \tag{3}$$

Since obtaining a global solution to this problem is still a difficult task in the general case, we analyze in this paper the probabilistic properties of any suboptimal solution. Furthermore, if at most m violations of the N constraints are allowed, the following sampled problem can be used to obtain a probabilistic relaxation to the original problem (2)

$$\min_{\theta \in \Theta} J(\theta) \;\text{ subject to } \sum_{\ell=1}^{N} g(\theta, w^{(\ell)}) \le m. \tag{4}$$

Randomized strategies to solve problems (3) and (4) have been recently studied in [2], see also [31]. In order to analyze the probabilistic properties of any feasible solution to problem (4), we introduce the definition of probability of failure.

Definition 2 *[probability of failure] Given* $N, \eta \in (0, 1)$, *the integer* m *where* $0 \le m \le N$ *and* $g : \Theta \times \mathcal{W} \to \{0, 1\}$, *the probability of failure, denoted by* $p(N, \eta, m)$ *is defined as*

$$\begin{aligned} p(N, \eta, m) := \Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N \;:\; & \textit{There exists } \theta \in \Theta \\ & \textit{such that } \hat{E}(\theta, \mathrm{w}) \le \frac{m}{N} \textit{ and } E(\theta) > \eta\}, \end{aligned}$$

which is slightly different than the probability of one-sided constrained failure introduced in [2]. We notice that if the probability of failure $p(N, \eta, m)$ is no greater than δ then every feasible solution $\theta \in \Theta$ to problem (4) satisfies $E(\theta) \le \eta$ with probability no smaller than $1 - \delta$. Therefore, the objective is to obtain explicit expressions yielding a minimum number of samples N such that the inequality $p(N, \eta, m) \le \delta$ holds.

We notice that the probability of failure can be easily bounded by the binomial distribution if $\Theta = \{\hat{\theta}\}$ is a singleton, which represents an analysis problem. In this case, $p(N, \eta, m)$ can be easily obtained as follows

$$p(N, \eta, m) =$$

$$\Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N \;:\; \hat{E}(\hat{\theta}, \mathrm{w}) \le \frac{m}{N} \text{ and } E(\hat{\theta}) > \eta\} =$$

$$\Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N \;:\; \sum_{\ell=1}^{N} g(\hat{\theta}, w^{(\ell)}) \le m \text{ and } E(\hat{\theta}) > \eta\} \le$$

$$\Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N \;:\; \sum_{\ell=1}^{N} g(\hat{\theta}, w^{(\ell)}) \le m \text{ and } E(\hat{\theta}) = \eta\} =$$

$$\sum_{i=0}^{m}\begin{pmatrix}N\\ i\end{pmatrix}\eta^i(1-\eta)^{N-i}.$$

In Subsection 4.2 we derive bounds on $p(N,\eta,m)$ when Θ consists of a finite number of elements. On the other hand, if Θ consists of an infinite number of elements, a deeper analysis involving Statistical Learning Theory is needed [31], [33]. In Subsection 4.3 we study the probabilistic properties of the optimal solution of problem (3) under the assumption that $g(\theta,w)=0$ is equivalent to $f(\theta,w)\leq 0$, where $f:\Theta\times\mathcal{W}\to\mathbb{R}$ is a convex function with respect to θ in Θ. In this case the result is not expressed in terms of probability of failure because it applies only to the optimal solution of problem (3), and not to every feasible solution.

3 Sample complexity for the binomial distribution

Given a positive integer N and a nonnegative integer m, $m\leq N$, and $\eta\in(0,1)$, the binomial distribution is given by

$$B(N,\eta,m):=\sum_{i=0}^{m}\begin{pmatrix}N\\ i\end{pmatrix}\eta^i(1-\eta)^{N-i}.$$

The problem we address in this section is the explicit computation of the *sample complexity*, i.e. a function $\tilde{N}(\eta,m,\delta)$ such that the inequality $B(N,\eta,m)\leq\delta$ holds for any $N\geq\tilde{N}(\eta,m,\delta)$, where $\delta\in(0,1)$. As it will be illustrated in the following section, the inequality $B(N,\eta,m)\leq\delta$ plays a fundamental role in probabilistic analysis and design methods. Although some explicit expressions are available, e.g. the multiplicative and additive forms of Chernoff bound [16], the results obtained in this paper are tuned on the specific inequalities stemming from the problems described in Section 4.

The following technical lemma provides an upper bound for the binomial distribution $B(N,\eta,m)$.

Lemma 1 *Suppose that $\eta\in(0,1)$ and that the nonnegative integer m and the positive integer N satisfy $m\leq N$. Then,*

$$\begin{aligned}B(N,\eta,m)&=\sum_{i=0}^{m}\begin{pmatrix}N\\ i\end{pmatrix}\eta^i(1-\eta)^{N-i}\\ &\leq a^m\left(\frac{\eta}{a}+1-\eta\right)^N,\ \ \forall a\geq 1.\end{aligned}$$

Proof:

The proof of the lemma follows from the following sequence of inequalities:

$$\begin{aligned}B(N,\eta,m)&=a^m\sum_{i=0}^{m}\begin{pmatrix}N\\ i\end{pmatrix}a^{-m}\eta^i(1-\eta)^{N-i}\\ &\leq a^m\sum_{i=0}^{m}\begin{pmatrix}N\\ i\end{pmatrix}a^{-i}\eta^i(1-\eta)^{N-i}\\ &\leq a^m\sum_{i=0}^{N}\begin{pmatrix}N\\ i\end{pmatrix}\left(\frac{\eta}{a}\right)^i(1-\eta)^{N-i}\\ &=a^m\left(\frac{\eta}{a}+1-\eta\right)^N.\end{aligned}$$

□

We notice that each particular choice of $a\geq 1$ provides an upper bound for $B(N,\eta,m)$. When using Lemma 1 to obtain a specific sample complexity, the selected value for a plays a significant role.

Lemma 2 *Given $\delta\in(0,1)$ and the nonnegative integer m, suppose that the integer N and the scalars $\eta\in(0,1)$ and $a>1$ satisfy the inequality*

$$N\geq\frac{1}{\eta}\left(\frac{a}{a-1}\right)\left(\ln\frac{1}{\delta}+m\ln a\right).\tag{5}$$

Then, $m\leq N$ and

$$B(N,\eta,m)=\sum_{i=0}^{m}\begin{pmatrix}N\\ i\end{pmatrix}\eta^i(1-\eta)^{N-i}\leq\delta.$$

Proof:

We first prove that if inequality (5) is satisfied then $N\geq m$. Since $\eta\in(0,1)$ and $\delta\in(0,1)$, (5) implies

$$N\geq\left(\frac{a}{a-1}\ln a\right)m.$$

Next, we notice that

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{a}{a-1}\ln a\right)=\left(\frac{-1}{(a-1)^2}\right)\ln a+\frac{1}{a-1}.$$

Since $\ln a<a-1$ for every $a>1$, it follows that

$$\frac{\mathrm{d}}{\mathrm{d}a}\left(\frac{a}{a-1}\ln a\right)>\left(\frac{-1}{(a-1)^2}\right)(a-1)+\frac{1}{a-1}=0.$$

Using this fact, we conclude that $\frac{a}{(a-1)} \ln a$ is a strictly increasing function for $a > 1$. This means that

$$N \geq \left(\frac{a}{a-1} \ln a\right) m \geq \lim_{\hat{a}\to 1} \left(\frac{\hat{a}}{\hat{a}-1} \ln \hat{a}\right) m = m.$$

We now prove that (5) guarantees that $a^m(\frac{\eta}{a}+1-\eta)^N \leq \delta$. The inequality (5) can be rewritten as

$$N\eta \left(\frac{a-1}{a}\right) \geq \ln \frac{1}{\delta} + m \ln a. \tag{6}$$

Since $x \leq -\ln(1-x)$ for every $x \in (0,1)$, and $\eta(\frac{a-1}{a}) \in (0,1)$, from inequality (6), we obtain a sequence of inequalities

$$-N \ln \left(1 - \eta \left(\frac{a-1}{a}\right)\right) \geq \ln \frac{1}{\delta} + m \ln a$$

$$\ln \delta \geq m \ln a + N \ln \left(1 - \eta \left(\frac{a-1}{a}\right)\right)$$

$$\delta \geq a^m \left(\frac{\eta}{a} + 1 - \eta\right)^N.$$

We have therefore proved that inequality (5) implies $m \leq N$ and $a^m(\frac{\eta}{a} + 1 - \eta)^N \leq \delta$. The claim of the property follows directly from Lemma 1. □

Obviously, the best sample size bound is obtained taking the infimum with respect to $a > 1$. However, this required to solve a one-dimensional optimization problem for given η, δ and m. We observe that a suboptimal value can be immediately obtained setting a equal to the Euler constant, which yields the sample complexity

$$N \geq \frac{1}{\eta} \left(\frac{\mathrm{e}}{\mathrm{e}-1}\right) \left(\ln \frac{1}{\delta} + m\right).$$

Since $\frac{\mathrm{e}}{\mathrm{e}-1} < 1.59$, we obtain $N \geq \frac{1.59}{\eta}\left(\ln \frac{1}{\delta} + m\right)$. We also notice that, if $m > 0$ then the choice $a = 1 + \frac{\ln \frac{1}{\delta}}{m} + \sqrt{2\frac{\ln \frac{1}{\delta}}{m}}$ provides a less conservative bound at the price of a more involved expression. Based on extensive numerical computations for several values of η, δ and m it seems that this bound is very close to the "optimal" one. The explicit bound presented in the following Corollary constitutes one of the main results of this paper.

Corollary 1 *Given $\delta \in (0,1)$ and the nonnegative integer m, suppose that the integer N and the scalar $\eta \in (0,1)$ satisfy the inequality*

$$N \geq \frac{1}{\eta} \left(m + \ln \frac{1}{\delta} + \sqrt{2m \ln \frac{1}{\delta}}\right). \tag{7}$$

Then,

$$B(N,\eta,m) = \sum_{i=0}^{m} \binom{N}{i} \eta^i (1-\eta)^{N-i} \leq \delta. \tag{8}$$

The proof of this corollary is shown in the appendix. This corollary improves upon the explicit expression obtained using the multiplicative form of the Chernoff bound [31], which is given by

$$N \geq \frac{1}{\eta} \left(m + \ln \frac{1}{\delta} + \sqrt{\left(\ln \frac{1}{\delta}\right)^2 + 2m \ln \frac{1}{\delta}}\right).$$

4 Sample complexity for probabilistic analysis and design

We now study some problems in the context of randomized algorithms where one encounters inequalities of the form

$$B(N,\eta,m) = \sum_{i=0}^{m} \binom{N}{i} \eta^i (1-\eta)^{N-i} \leq \delta.$$

In particular, we show how the results of the previous section can be used to obtain explicit sample size bounds which guarantee that the probabilistic solutions obtained from different randomized approaches meet some pre-specified probabilistic properties.

4.1 Worst-case performance analysis

We recall a result shown in [30] for the probabilistic worst-case performance analysis.

Theorem 1 *Assume that, given the function $f : \Theta \times \mathcal{W} \to \mathbb{R}$, and $\hat{\theta} \in \Theta$, the multisample $\mathrm{w} = \{w^{(1)}, \ldots, w^{(N)}\}$ is drawn from $\mathcal{W}^N$ according to probability $\Pr_{\mathcal{W}^N}$. Suppose also that*

$$\gamma = \max_{\ell=1,\ldots,N} f(\hat{\theta}, w^{(\ell)}).$$

If

$$N \geq \frac{\ln \frac{1}{\delta}}{\ln \frac{1}{1-\eta}},$$

then $\Pr_{\mathcal{W}}\{w \in \mathcal{W} : f(\hat{\theta}, w) > \gamma\} \leq \eta$ with probability no smaller than $1-\delta$.

The proof of this statement can be found in [30] and is based on the fact that $\Pr_{\mathcal{W}}\{w \in \mathcal{W} : f(\hat{\theta}, w) > \gamma\} \leq \eta$ with probability no smaller than $1-(1-\eta)^N$. Therefore, it suffices to take N such that $B(N,\eta,0) = (1-\eta)^N \leq \delta$.

4.2 Finite families for design

We consider the non-convex sampled problem (4) for the case when Θ consists of a set of finite cardinality n_C. As a motivation, we study the case when, after an appropriate normalization procedure, the design parameter set is rewritten as $\hat{\Theta} = \{ \ \theta \in \mathbb{R}^{n_\theta} \ : \ \|\theta\|_\infty \leq 1 \ \}$. Suppose also that a gridding approach is adopted. That is, for each component θ_j, $j = 1, \ldots, n_\theta$ of the design parameters $\theta \in \mathbb{R}^{n_\theta}$, only n_{C_j} equally spaced values are considered. That is, θ_j is constrained into the set $\Upsilon_j = \{ \ -1 + \frac{2(t-1)}{(n_{C_j}-1)} \ : \ t = 1, \ldots, n_{C_j} \ \}$. With this gridding procedure, the following finite cardinality set $\Theta = \{ \ [\theta_1, \ldots, \theta_{n_\theta}]^T \ : \ \theta_j \in \Upsilon_j, \ j = 1, \ldots, n_\theta \ \}$ is obtained. We notice that the cardinality of the set is $n_C = \prod_{j=1}^{n_\theta} n_{C_j}$. Another situation in which the finite cardinality assumption holds is when a finite number of random samples in the space of design parameter are drawn according to a given probability, see e.g. [18], [33].

The following property states the relation between the binomial distribution and the probability of failure under this finite cardinality assumption.

Lemma 3 *Suppose that the cardinality of Θ is no larger than n_C. Then,*

$$p(N, \eta, m) \leq n_C \sum_{i=0}^{m} \binom{N}{i} \eta^i (1-\eta)^{N-i} = n_C B(N, \eta, m).$$

Proof:

Denote $\tilde{n}_C \leq n_C$ the cardinality of Θ. Therefore, Θ can be rewritten as $\Theta = \{\theta^{(1)}, \theta^{(2)}, \ldots, \theta^{(\tilde{n}_C)}\}$. Then, the following inequalities can be easily obtained.

$$\begin{aligned} p(N, \eta, m) = \Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N \ : \ \text{There exists } \theta \in \Theta \\ \text{such that } \hat{E}(\theta, \mathrm{w}) \leq \frac{m}{N} \text{ and } E(\theta) > \eta\} \leq \\ \sum_{j=1}^{\tilde{n}_C} \Pr_{\mathcal{W}^N}\{\mathrm{w} \in \mathcal{W}^N : \hat{E}(\theta^{(j)}, \mathrm{w}) \leq \\ \frac{m}{N} \text{ and } E(\theta^{(j)}) > \eta\} \leq \tilde{n}_C \sum_{i=0}^{m} \binom{N}{i} \eta^i (1-\eta)^{N-i} \leq \\ n_C \sum_{i=0}^{m} \binom{N}{i} \eta^i (1-\eta)^{N-i}. \end{aligned}$$

□

Consider now the optimization problem (4). It follows from Lemma 3 that to guarantee that every feasible solution $\hat{\theta} \in \Theta$ satisfies $E(\hat{\theta}) \leq \eta$ with probability no smaller than $1-\delta$, it suffices to take N such that $n_C B(N, \eta, m) \leq \delta$, where n_C is an upper bound on the cardinality of Θ. As it will be shown next, the required sample complexity in this case grows with the logarithm of n_C. This means that we can consider finite families with high cardinality and still obtain very reasonable sample complexity bounds.

Theorem 2 *Suppose that the cardinality of Θ is no larger than n_C. Given the nonnegative integer m, $\eta \in (0,1)$ and $\delta \in (0,1)$, if*

$$N \geq \inf_{a>1} \frac{1}{\eta} \left(\frac{a}{a-1}\right) \left(\ln \frac{n_C}{\delta} + m \ln a\right) \tag{9}$$

then $p(N, \eta, m) \leq \delta$. Moreover, if

$$N \geq \frac{1}{\eta} \left(m + \ln \frac{n_C}{\delta} + \sqrt{2m \ln \frac{n_C}{\delta}}\right)$$

then $p(N, \eta, m) \leq \delta$.

Proof:

From Lemma 3 we have that $p(N, \eta, m) \leq \delta$ provided that $B(N, \eta, m) \leq \frac{\delta}{n_C}$. The two claims of the property now follow directly from Lemma 2 and Corollary 1 respectively.

□

Consider the sampled optimization problem (4)

$$\min_{\theta \in \Theta} J(\theta) \ \text{ subject to } \sum_{\ell=1}^{N} g(\theta, w^{(\ell)}) \leq m.$$

From the definition of $p(N, \eta, m)$ and Theorem 2 we conclude that if one draws N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$, then with probability no smaller than $1-\delta$, all the feasible solutions to problem (4) have a probability of violation no larger than η, provided that the cardinality of Θ is upper bounded by n_C and the sample complexity is given by

$$N \geq \frac{1}{\eta} \left(m + \ln \frac{n_C}{\delta} + \sqrt{2m \ln \frac{n_C}{\delta}}\right).$$

We remark that taking a equal to the Euler constant, the following sample size bound

$$N \geq \frac{1}{\eta} \left(\frac{\mathrm{e}}{\mathrm{e}-1}\right) \left(\ln \frac{n_C}{\delta} + m\right)$$

is immediately obtained from Theorem 2. If $m > 0$ then a suboptimal value for a is given by

$$a = 1 + \frac{\ln \frac{n_C}{\delta}}{m} + \sqrt{2\frac{\ln \frac{n_C}{\delta}}{m}}.$$

4.3 Optimal robust optimization for design

In this subsection, we study the so-called scenario approach for robust control introduced in [10]. To address the semi-infinite optimization problem (2), we solve the sampled optimization problem (3). That is, we generate N i.i.d. samples $\{w^{(1)}, \ldots, w^{(N)}\}$ from $\mathcal{W}$ according to the probability $\Pr_{\mathcal{W}}$ and then solve the following problem:

$$\min_{\theta \in \Theta} J(\theta) \;\text{ subject to } g(\theta, w^{(\ell)}) = 0,\; \ell = 1, \ldots, N. \quad (10)$$

We consider here the particular case in which $J(\theta) = c^T\theta$, the constraint $g(\theta, w) = 0$ is convex in θ for all $w \in W$ and the solution of (10) is unique. These assumptions are now stated precisely.

Assumption 1 [convexity] *Let $\Theta \subset \mathbb{R}^{n_\theta}$ be a convex and closed set. We assume that*

$$J(\theta) := c^T\theta \quad \textit{and} \quad g(\theta, w) := \begin{cases} 0 & \textit{if } f(\theta, w) \le 0, \\ 1 & \textit{otherwise} \end{cases}$$

where $f : \Theta \times \mathcal{W} \to [-\infty, \infty]$ is convex in θ for every fixed value of $w \in \mathcal{W}$.

Assumption 2 [feasibility and uniqueness] *For all possible multisample extractions $\{w^{(1)}, \ldots, w^{(N)}\}$, the optimization problem (10) is always feasible and attains a unique optimal solution. Moreover, its feasibility domain has a nonempty interior.*

We now state a result proved in [14], [10] that relates the binomial distribution to the probabilistic properties of the optimal solution obtained from (10).

Lemma 4 *Let Assumptions 1 and 2 hold. Suppose that N, $\eta \in (0, 1)$ and $\delta \in (0, 1)$ satisfy the inequality*

$$\sum_{i=0}^{n_\theta - 1} \binom{N}{i} \eta^i (1-\eta)^{N-i} \le \delta. \quad (11)$$

Then, with probability no smaller than $1 - \delta$, the optimal solution $\hat{\theta}_N$ to the optimization problem (10) satisfies the inequality $E(\hat{\theta}_N) \le \eta$.

We remark that the assumptions stated in Lemma 4 can be relaxed. Moreover, this result has been extended to the possible unfeasible case, see [12], [14] and [15].

We now state an explicit sample size bound, which improves upon previous bounds, to guarantee that the probability of violation is smaller than η with probability at least $1 - \delta$.

Theorem 3 *Let Assumptions 1 and 2 hold. Given $\eta \in (0, 1)$ and $\delta \in (0, 1)$, if*

$$N \ge \inf_{a > 1} \left(\frac{a}{\eta(a-1)} \right) \left(\ln \frac{1}{\delta} + (n_\theta - 1) \ln a \right) \quad (12)$$

or

$$N \ge \frac{1}{\eta} \left(\ln \left(\frac{1}{\delta} \right) + n_\theta - 1 + \sqrt{2(n_\theta - 1) \ln \frac{1}{\delta}} \right) \quad (13)$$

then, with probability no smaller than $1 - \delta$, the optimal solution $\hat{\theta}_N$ to the optimization problem (10) satisfies the inequality $E(\hat{\theta}_N) \le \eta$.

Proof: From Lemma 4 it follows that it suffices to take N such that $B(N, \eta, n_\theta - 1) \le \delta$. Both inequalities (12) and (13) guarantee that $B(N, \eta, n_\theta - 1) \le \delta$ (see Lemma 2 and Corollary 1 respectively). This completes the proof. □

We remark that a sample size bound which depends linearly on $\frac{1}{\eta}$ is simply obtained taking a equal to the Euler constant

$$N \ge \frac{1}{\eta} \left(\frac{\mathrm{e}}{\mathrm{e} - 1} \right) \left(\ln \frac{1}{\delta} + n_\theta - 1 \right).$$

If $n_\theta > 1$ then a suboptimal value for a is given by

$$a = 1 + \frac{\ln \frac{1}{\delta}}{n_\theta - 1} + \sqrt{2\frac{\ln \frac{1}{\delta}}{n_\theta - 1}}.$$

5 Sequential algorithms with probabilistic validation

In this section we present a general family of randomized algorithms, which we denote as "Sequential Probabilistic Validation (SPV) algorithms". The main feature of this class of algorithms is that they are based on a probabilistic validation step. This family includes most of the sequential randomized algorithms that have been discussed in the introduction of this paper. As a matter of fact, the non-sequential strategies that can be found in the context of Statistical Learning Theory [33] and convex scenario [10] can be also provided with an outer

iterative structure that makes them fit in the proposed scheme. We explore this possibility in Section 8.

Each iteration of an SPV algorithm includes the computation of a candidate solution for the problem and a validation step. The results provided in this paper are basically independent of the particular strategy chosen to obtain candidate solutions. Therefore, in the following discussion we restrict ourselves to a generic candidate solution. The accuracy $\eta \in (0,1)$ and confidence $\delta \in (0,1)$ required for the probabilistic solution play a relevant role when determining the sample size of each validation step. The main purpose of this part of the paper is to provide a validation scheme such that it guarantees that for given accuracy η and confidence δ, all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no larger than η with probability no smaller than $1-\delta$.

We enumerate each iteration of the algorithm by means of an integer k. We denote by m_k the number of violations that are allowed at the validation step of iteration k. We assume that m_k is given by a function of k, that is, $m_k = m(k)$ where the function $m : \mathbb{N} \to \mathbb{N}$ is given. We also denote by M_k the sample size of the validation step of iteration k. We assume that M_k is given by a function of k, η and δ. That is, $M_k = M(k, \eta, \delta)$ where $M : \mathbb{N} \times \mathbb{R} \times \mathbb{R} \to \mathbb{N}$ has to be appropriately designed in order to guarantee the probabilistic properties of the algorithm. In fact, one of the main contributions of [24] is to provide this function for the particular case $m_k = 0$ for every $k \geq 1$. The functions $m(\cdot)$ and $M(\cdot,\cdot,\cdot)$ are denoted as *level function* and *cardinality function* respectively.

We now introduce the structure of an SPV algorithm

(i) Set accuracy $\eta \in (0,1)$ and confidence $\delta \in (0,1)$ equal to the desired levels. Set k equal to 1.
(ii) Obtain a candidate solution $\hat{\theta}_k$ to the robust optimization problem (1).
(iii) Set $m_k = m(k)$ and $M_k = M(k, \eta, \delta)$.
(iv) Obtain validation set $\mathcal{V}_k = \{v^{(1)}, \ldots, v^{(M_k)}\}$ drawing M_k i.i.d validation samples from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$.
(v) If $\sum\limits_{\ell=1}^{M_k} g(\hat{\theta}_k, v^{(\ell)}) \leq m_k$, then $\hat{\theta}_k$ is a probabilistic solution.
(vi) Exit if the exit condition is satisfied.
(vii) $k = k+1$. Goto (ii).

Although the exit condition can be quite general, a reasonable choice is to exit after a given number of candidate solutions have been classified as probabilistic solutions or when a given computational time has elapsed since the starting of the algorithm. After exiting one could choose the probabilistic solution which maximizes a given performance index. In the next section we propose a strategy to choose the cardinality of the validation set at iteration k in such a way that with probability no smaller than $1-\delta$ all candidate solutions classified as probabilistic solutions by the algorithm meet the accuracy η.

6 Adjusting the validation sample size

The cardinality adjusting strategy provided in this section constitutes a generalization of that presented in [24] and [17]. To obtain the results of this section we rely on some contributions on the sample complexity presented in the previous sections.

We now formally introduce the *failure function*.

Definition 3 (failure function) *The function $\mu : \mathbb{N} \to \mathbb{R}$ is said to be a failure function if it satisfies the following conditions:*

(i) $\mu(k) \in (0,1)$ *for every positive integer k.*
(ii) $\sum\limits_{k=1}^{\infty} \mu(k) \leq 1$.

We notice that the function

$$\mu(k) = \frac{1}{\xi(\alpha)k^{\alpha}},$$

where $\xi(\cdot)$ is the Riemann zeta function, is a failure function for every $\alpha > 1$. This is due to the fact that $\sum\limits_{i=1}^{\infty} \frac{1}{k^{\alpha}}$ converges for every scalar α greater than 1 to $\xi(\alpha)$. This family has been used in the context of validation schemes in [13] and in [24] for the particular value $\alpha = 2$.

Property 1 *Consider an SPV algorithm with given accuracy parameter $\eta \in (0,1)$, confidence $\delta \in (0,1)$, level function $m(\cdot)$ and cardinality function $M(\cdot,\cdot,\cdot)$. If there exists a failure function $\mu(\cdot)$ such that*

$$\sum_{i=0}^{m(k)} \begin{pmatrix} M(k,\eta,\delta) \\ i \end{pmatrix} \eta^i (1-\eta)^{M(k,\eta,\delta)-i} \leq \delta\mu(k), \quad \forall k \geq 1$$

then with probability greater than $1-\delta$ all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no greater than η.

The following proof of this property follows the same lines as the proof of Theorem 9 in [24].

Proof:

We denote by δ_k the probability of classifying at iteration k the candidate solution $\hat{\theta}_k$ as a probabilistic solution

under the assumption that the probability of violation $E(\hat{\theta}_k)$ is larger than η. Then

$$\begin{aligned}
\delta_k &= \mathrm{Pr}_{\mathcal{W}^{M_k}} \Big\{ \{v^{(1)}, \ldots, v^{(M_k)}\} \in \mathcal{W}^{M_k} \ : \\
&\quad \sum_{j=1}^{M_k} g(\hat{\theta}_k, v^{(j)}) \leq m_k \ \text{ and } E(\hat{\theta}_k) > \eta \Big\} \\
&< \mathrm{Pr}_{\mathcal{W}^{M_k}} \Big\{ \{v^{(1)}, \ldots, v^{(M_k)}\} \in \mathcal{W}^{M_k} \ : \\
&\quad \sum_{j=1}^{M_k} g(\hat{\theta}_k, v^{(j)}) \leq m_k \ \text{ and } E(\hat{\theta}_k) = \eta \Big\} \\
&= \sum_{i=0}^{m_k} \begin{pmatrix} M_k \\ i \end{pmatrix} \eta^i (1-\eta)^{M_k - i} \\
&= \sum_{i=0}^{m(k)} \begin{pmatrix} M(k,\eta,\delta) \\ i \end{pmatrix} \eta^i (1-\eta)^{M(k,\eta,\delta) - i} \leq \delta\mu(k).
\end{aligned}$$

Therefore, the probability of misclassification of a candidate solution at iteration k is smaller than $\delta\mu(k)$. We conclude that the probability of erroneously classifying one or more candidate solutions as probabilistic solutions is bounded by

$$\sum_{k=1}^{\infty} \delta_k < \sum_{k=1}^{\infty} \delta\mu(k) = \delta \sum_{k=1}^{\infty} \mu(k) \leq \delta.$$

□

To design a cardinality function $M(\cdot,\cdot,\cdot)$ satisfying the conditions of Property 1 we use Corollary 1.

We now present the main contribution of this part of the paper, which is a general expression for the cardinality of the validation set at each iteration of the algorithm.

Theorem 4 *Consider an SPV algorithm with given accuracy $\eta \in (0,1)$, confidence $\delta \in (0,1)$ and level function $m(\cdot)$. Suppose also that $\mu(\cdot)$ is a failure function. Then, the cardinality function*

$$M(k,\eta,\delta) =$$

$$\left\lceil \frac{1}{\eta} \left(m(k) + \ln \frac{1}{\delta\mu(k)} + \sqrt{2m(k) \ln \frac{1}{\delta\mu(k)}} \right) \right\rceil$$

guarantees that, with probability greater than $1-\delta$, all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no greater than η.

Proof:

Corollary 1 guarantees that the proposed choice for the cardinality function satisfies the inequality

$$\sum_{i=0}^{m(k)} \begin{pmatrix} M(k,\eta,\delta) \\ i \end{pmatrix} \eta^i (1-\eta)^{M(k,\eta,\delta) - i} \leq \delta\mu(k), \ \forall k \geq 1.$$

The result then follows from a direct application of Property 1. □

We notice that the proposed cardinality function $M(k,\eta,\delta)$ in Theorem 4 depends on the previous selection of the level function $m(\cdot)$ and the failure function $\mu(\cdot)$. Reasonable choices for these functions are $m(k) = \lfloor ak \rfloor$, where a is a non-negative scalar and $\mu(k) = \frac{1}{\xi(\alpha)k^\alpha}$ where α is greater than one. We recall that this choice guarantees that $\mu(k)$ is a failure function. As shown in the following section, the proposed level and failure functions allow us to recover, for the particular choice $a = 0$ the validation strategies proposed in [17] and [24]. In the next corollary, we rewrite the generic structure of the SPV algorithm with the level function $m(k) = \lfloor ak \rfloor$, and state a probabilistic result.

Corollary 2 *Consider the following SPV algorithm*

(i) Set accuracy $\eta \in (0,1)$, confidence $\delta \in (0,1)$ and scalars $a \geq 0$, $\alpha > 1$ equal to the desired levels. Set k equal to 1.
(ii) Obtain a candidate solution $\hat{\theta}_k$ to the robust optimization problem (1).
(iii) Set $m_k = \lfloor ak \rfloor$ and

$$M_k = \left\lceil \frac{1}{\eta} \left(m_k + \ln \frac{\xi(\alpha)k^\alpha}{\delta} + \sqrt{2m_k \ln \frac{\xi(\alpha)k^\alpha}{\delta}} \right) \right\rceil.$$

(iv) Obtain the validation set $\mathcal{V}_k = \{v^{(1)}, \ldots, v^{(M_k)}\}$ drawing M_k i.i.d. validation samples from $\mathcal{W}$ according to the probability $\mathrm{Pr}_{\mathcal{W}}$.
(v) If $\sum_{\ell=1}^{M_k} g(\hat{\theta}_k, v^{(\ell)}) \leq m_k$, then $\hat{\theta}_k$ is a probabilistic solution.
(vi) Exit if the exit condition is satisfied.
(vii) k=k+1. Goto (ii).

Then, with probability greater than $1-\delta$ all the probabilistic solutions obtained running the SPV algorithm have a probability of violation no greater than η.

Proof: The result is obtained directly from Theorem 4 using as level function $m(k) = \lfloor ak \rfloor$ and failure function $\mu(k) = \frac{1}{\xi(\alpha)k^\alpha}$. □

Since the probabilistic properties of the algorithm presented in Corollary 2 are independent of the particular value of $\alpha > 1$, a reasonable choice for α is that minimizing the cardinality of the validation samples. In [17] it is shown that $\alpha = 1.1$ is an appropriate choice leading to a minimization of the cardinality of the validation sets.

7 Comparison with other validation schemes

In this section, we provide comparisons with the validation schemes presented in [24]. This strategy has been successfully used in different randomized algorithms dealing with uncertain convex problems e.g. in [1], [11], [24]. We notice that setting $a = 0$ and $\alpha = 2$ in Corollary 2 we obtain $m(k) = 0$ for every iteration k and

$$M(k) = \left\lceil \frac{1}{\eta} \ln \left(\frac{\xi(2)k^2}{\delta} \right) \right\rceil = \left\lceil \frac{1}{\eta} \ln \left(\frac{\pi^2 k^2}{6\delta} \right) \right\rceil.$$

This is the same cardinality function presented in [24] if one takes into account that for small values of η, $-\ln(1-\eta)$ can be approximated by η. In the same way, $a = 0$ and $\alpha = 1.1$ lead to the cardinality function presented in [17].

We noticed that not allowing any failure in each validation test makes perfect sense for convex problems if the feasibility set

$$\Theta_r = \{ \theta \in \Theta \ : \ g(\theta, w) = 0 \text{ for all } w \in \mathcal{W} \}$$

is not empty. Under this assumption, the algorithm takes advantage of the validation samples that have not satisfied the specifications to obtain a new candidate solution. If Θ_r is not empty, a common feature of the methods which use this strict validation scheme is that a probabilistic solution (not necessarily belonging to the feasibility set Θ_r) is obtained in a finite number of iterations of the algorithm, see e.g., [1], [11], [24].

A very different situation is encountered when Θ_r is empty. We now state a property showing that we should not use a strict validation scheme ($a = 0$) to address the case of empty robust feasible set because the algorithm might fail to obtain a probabilistic solution even if the set $\{ \theta \in \Theta \ : \ E(\theta) \leq \eta \}$ is not empty.

Property 2 *Consider the SPV algorithm presented in Corollary 2 with $a = 0$ and $\alpha > 1$. Suppose that $E(\theta) \geq \mu > 0$ for all $\theta \in \Theta$. Then the SPV algorithm does not find any probabilistic solution in the first N iterations of the algorithm with probability greater than*

$$1 - \left(\frac{\delta}{\xi(\alpha)} \right)^{\frac{\mu}{\eta}} \Phi(\frac{\alpha\mu}{\eta}, \lceil \log_2 N \rceil),$$

where the function $\Phi(s,t)$ is given by

$$\Phi(s,t) := \begin{cases} \dfrac{1 - 2^{(1-s)(t+1)}}{1 - 2^{1-s}} & \text{if } s \neq 1 \\ \\ t+1 & \text{otherwise} \end{cases}$$

and s is a strictly positive scalar and t is a non-negative integer.

Proof:

We notice that $a = 0$ implies that, at iteration k, the algorithm classifies a candidate solution $\hat{\theta}_k$ as a probabilistic solution only if it satisfies the constraint $g(\hat{\theta}_k, v^{(k)}) = 0, k = 1, \ldots, M_k$ where $\{v^{(1)}, \ldots, v^{(M_k)}\}$ is the randomly obtained validating set $\mathcal{V}_k$. Since $E(\theta) \geq \mu$ for all $\theta \in \Theta$ and $a = 0$, the probability of classifying a candidate solution as a probabilistic solution at iteration k is no greater than

$$(1-\mu)^{M_k} = \mathrm{e}^{M_k \ln(1-\mu)} < \mathrm{e}^{-\mu M_k}$$

$$\leq \mathrm{e}^{-\frac{\mu}{\eta} \ln \left(\frac{\xi(\alpha)k^\alpha}{\delta} \right)} = \left(\frac{\delta}{\xi(\alpha)k^\alpha} \right)^{\frac{\mu}{\eta}}.$$

Therefore, the probability of providing a probabilistic solution at any of the first N iterations of the algorithm is smaller than

$$\sum_{k=1}^{N} \left(\frac{\delta}{\xi(\alpha)k^\alpha} \right)^{\frac{\mu}{\eta}} = \left(\frac{\delta}{\xi(\alpha)} \right)^{\frac{\mu}{\eta}} \sum_{k=1}^{N} \left(\frac{1}{k^\alpha} \right)^{\frac{\mu}{\eta}}.$$

Taking $s = \frac{\alpha\mu}{\eta}$ and using Property 4 in the Appendix we have

$$\sum_{k=1}^{N} \left(\frac{1}{k^\alpha} \right)^{\frac{\mu}{\eta}} = \sum_{k=1}^{N} \frac{1}{k^s} \leq \Phi(s, \lceil \log_2 N \rceil).$$

We conclude that the probability of not finding any probabilistic solution in the first N iterations of the algorithm is smaller than

$$1 - \left(\frac{\delta}{\xi(\alpha)} \right)^{\frac{\mu}{\eta}} \Phi(\frac{\alpha\mu}{\eta}, \lceil \log_2 N \rceil).$$

$\square$

We now present an example demonstrating that a strict validation scheme may not be well-suited for a robust design problem.

Example 1 *Suppose that $\Theta = [0,1]$, $\mathcal{W} = [-0.08, 1]$, $\eta = 0.1$, $\delta = 10^{-4}$ and that*

$$g(\theta, w) = \begin{cases} 0 & \text{if } \theta \leq w \\ 1 & \text{otherwise.} \end{cases}$$

Suppose also that $\Pr_{\mathcal{W}}$ is the uniform distribution. It is clear that $\theta = 0$ minimizes the probability of violation and satisfies $E(0) = \frac{0.08}{1.08} > 0.074$. Therefore, we obtain

$$E(\theta) \geq 0.074 = \mu \text{ for all } \theta \in \Theta.$$

Consider now the choice $\alpha = 1.1$ and a maximum number of iterations N equal to 10^6. We conclude from Property 2 that, regardless of the strategy used to obtain candidate solutions, the choice $a = 0$ and $\alpha = 1.1$ in Corollary 2 does not find any probabilistic solution with probability greater than 0.98. The choice $\alpha = 2$ leads to a probability greater than 0.99. This illustrates that a strict validation scheme is not well suited for this robust design problem. □

The next result states that the validation scheme presented in this paper achieves, under minor technical assumptions, a solution with probability one.

Property 3 *Consider an SPV algorithm with given accuracy parameter $\eta \in (0,1)$, confidence $\delta \in (0,1)$ and level function $m(\cdot)$. Suppose that*

(i) $\mu(\cdot)$ is a failure function.
(ii) The cardinality function $M(k,\eta,\delta)$ is given by

$$\left\lceil \frac{1}{\eta}\left(m(k) + \ln\frac{1}{\delta\mu(k)} + \sqrt{2m(k)\ln\frac{1}{\delta\mu(k)}}\right)\right\rceil.$$

(iii) There exist an integer k^*, scalars $\mu \in (0,1)$ and $p \in (0,1)$ such that at every iteration $k > k^*$ a candidate solution $\hat{\theta}_k$ satisfying $E(\hat{\theta}_k) \leq \mu < \eta$ is obtained with probability greater than p.
(iv) $\lim\limits_{k\to\infty} \frac{1}{m(k)} \ln \frac{1}{\delta\mu(k)} = 0$.

Then the SPV algorithm achieves with probability 1 a solution in a finite number of iterations.

Proof:

Using the assumption

$$\lim_{k\to\infty} \frac{1}{m(k)} \ln \frac{1}{\delta\mu(k)} = 0$$

we conclude that

$$\lim_{k\to\infty} \frac{M(k)}{m(k)} =$$

$$\lim_{k\to\infty} \frac{1}{\eta}\left(1 + \frac{1}{m(k)}\ln\frac{1}{\delta\mu(k)} + \sqrt{2\frac{1}{m(k)}\ln\frac{1}{\delta\mu(k)}}\right) = \frac{1}{\eta}.$$

This implies, along with the assumption $\mu < \eta$, that there is $\tilde{k}$ such that

$$\mu < \frac{m(k)}{M(k)},\ \text{for every } k > \tilde{k}.$$

We conclude that the SPV algorithm provides candidate solutions $\hat{\theta}_k$ satisfying

$$E(\hat{\theta}_k) \leq \mu < \frac{m(k)}{M(k)} \tag{14}$$

for every $k \geq \max\{k^*, \tilde{k}\}$ with probability no smaller than p. The validation test is satisfied if

$$\sum_{\ell=1}^{M(k)} g(\hat{\theta}_k, v^{(\ell)}) \leq m(k),$$

or equivalently, if

$$\frac{1}{M(k)} \sum_{\ell=1}^{M(k)} g(\hat{\theta}_k, v^{(\ell)}) \leq \frac{m(k)}{M(k)}.$$

We notice that $\frac{1}{M(k)} \sum\limits_{\ell=1}^{M(k)} g(\hat{\theta}_k, v^{(\ell)})$ is the empirical mean associated to $g(\hat{\theta}_k, v)$. Moreover, recall that the probability of obtaining an empirical mean greater than the actual probability of violation is smaller than $\frac{1}{2}$ [31]. We therefore conclude from (14) that the probability of classifying a candidate solution as a probabilistic one is no smaller than $\frac{p}{2}$ for every iteration $k > \max\{k^*, \tilde{k}\}$. Since $\frac{p}{2} > 0$ we conclude that the algorithm obtains a probabilistic solution with probability 1. □

8 Application to non-sequential randomized algorithms

Motivated by the previous results we present an SPV algorithm which takes advantage of the theoretical results obtained in the literature of non-sequential randomized algorithms, see e.g.,[2].

(i) Set accuracy $\eta \in (0,1)$, confidence $\delta \in (0,1)$, and scalars $a > 0$ and $\alpha > 1$, equal to the desired levels.
(ii) Set k equal to 1 and $\mathcal{W}_1$ equal to the empty set.
(iii) Set $m_k = \lfloor ak \rfloor$ and

$$M_k = \left\lceil \frac{1}{\eta}\left(m_k + \ln\frac{\xi(\alpha)k^\alpha}{\delta} + \sqrt{2m_k \ln\frac{\xi(\alpha)k^\alpha}{\delta}}\right)\right\rceil.$$

(iv) Obtain, if possible, a candidate suboptimal feasible solution $\hat{\theta}_k$ to the optimization problem

$$\min_{\theta\in\Theta} J(\theta) \ \text{ subject to } \sum_{w\in\mathcal{W}_k} g(\theta, w) \leq (\operatorname{card} \mathcal{W}_k)\frac{m_k}{M_k}.$$

(v) Obtain validation set $\mathcal{V}_k = \{v^{(1)}, \ldots, v^{(M_k)}\}$ drawing M_k i.i.d validation samples from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$.

(vi) If a feasible solution $\hat{\theta}_k$ was found at step (iv) then classify it as a probabilistic solution if

$$\sum_{\ell=1}^{M_k} g(\hat{\theta}_k, v^{(\ell)}) \leq m_k.$$

(vii) Exit if the exit condition is satisfied.
(viii) $\mathcal{W}_{k+1} = \mathcal{W}_k \bigcup \mathcal{V}_k$, $k = k+1$. Goto (iii).

Under rather general assumptions like finite VC-dimension of the binary function $g(\cdot,\cdot)$, see [2] for further details, it follows that the feasible solutions obtained at step (iii) of the algorithm have a probability of violation smaller than $\frac{m_k}{M_k} < \eta$, with a probability that tends to 1 with the cardinality of $\mathcal{W}_k$. From this observation we conclude that if the optimization problem at step (iii) is feasible with probability greater than zero, the proposed algorithm satisfies the assumptions of Property 3 and therefore provides a probabilistic solution that meets accuracy η and confidence δ. The main advantage of the proposed algorithm with respect to the non-sequential algorithms available in the literature is that no explicit bound on the number of samples is required. This might lead to a substantial reduction of the number of required samples in particular in the case of non-convex uncertain problems.

9 Numerical example

The objective of this numerical example is to obtain probabilistic upper and lower bounds of a given time function $y : \mathcal{W} \to \mathbb{R}$ of the form

$$y(w) = [A(1+\frac{1}{2}t^2)\sin(7t+0.5)+B]\mathrm{e}^{-\frac{3}{2}t},$$

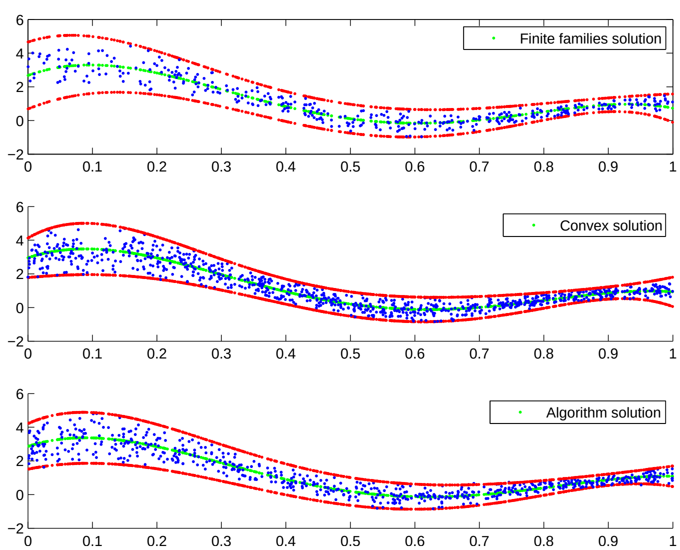

Fig. 1. Initial data set and envelope of the set of solutions.

where $w \in \mathcal{W}$.

The uncertainty set $\mathcal{W}$ is

$$\mathcal{W} = \left\{ w = [t \quad A \quad B]^T, t \in [0,1],\ A \in [1,3],\ B \in [1,3] \right\}.$$

For a given order d, we define the regressor $\varphi_d : \mathcal{W} \to \mathbb{R}^{d+1}$ as

$$\varphi_d(w) = \varphi_d([t \quad A \quad B]^T) = \left[1\ t\ t^2\ \cdots\ t^d \right]^T.$$

The objective of this example is to find a parameter vector $\theta = [\gamma_d, \lambda_d]^T$, $\gamma_d \in \mathbb{R}^{d+1}$ and $\lambda_d \in \mathbb{R}^{d+1}$ such that, with probability no smaller than $1-\delta$,

$$\mathrm{Pr}_{\mathcal{W}}\{\ w \in \mathcal{W}\ :\ |y(w) - \gamma_d^T \varphi_d(w)| \geq \lambda_d^T |\varphi_d(w)|\ \} \leq \eta.$$

The vector $|\varphi_d(w)|$ is obtained from the absolute values of $\varphi_d(w)$. The binary function $g : \Theta \times \mathcal{W} \to \{0,1\}$, is defined as

$$g(\theta, w) := \begin{cases} 0 & \text{if } \theta \text{ meets design specifications for } w \\ 1 & \text{otherwise,} \end{cases}$$

where "design specifications" means satisfying the following constraint:

$$|y(w) - \gamma_d^T \varphi_d(w)| \leq \lambda_d^T |\varphi_d(w)|$$

for randomly generated samples $w \in \mathcal{W}$.

A similar problem is addressed in [14] using the scenario approach. For the numerical computations, we take $\delta = 10^{-6}$ and $\eta = 0.05$. We address the problem from the finite families, scenario and SPV approach, and use the explicit sample complexity derived in the previous sections.

9.1 *Finite families approach*

We apply the results of Section 4.2 to determine both the degree d and the parameter vectors (γ_d, λ_d) that meet the design specification and optimize a given performance index.

In this example, as it will be seen later, a finite family of cardinality $n_C = 150$ is considered. In order to compare the finite family approach with the scenario one, we consider no allowed failures (i.e $m = 0$). For this choice of parameters ($m = 0$, $n_C = 150$, $\delta = 10^{-6}$ and $\eta = 0.05$), the number of samples N required to obtain a solution with the specified probabilistic probabilities is 377 (see Property 2). A set D of $M = N$ samples is drawn (i.i.d.)

from $\mathcal{W}$. We use these samples to select the optimal parameters $(\tilde{\gamma}_d, \tilde{\lambda}_d)$ corresponding to each of the different regressors $\varphi_d(\cdot)$. Each pair $(\tilde{\gamma}_d, \tilde{\lambda}_d)$ is obtained minimizing the empirical mean of the absolute value of the approximation error. That is, each pair $(\tilde{\gamma}_d, \tilde{\lambda}_d)$ is the solution to the optimization problem

$$\begin{aligned} \min_{\gamma_d, \lambda_d} \quad & \frac{1}{M} \sum_{w \in D} \lambda_d^T |\varphi_d(w)| \\ s.t. \quad & |y(w) - \gamma_d^T \varphi_d(w)| \leq \lambda_d^T |\varphi_d(w)|, \ \forall w \in D. \end{aligned}$$

We notice that the obtained parameters do not necessarily satisfy the probabilistic design specifications. In order to resolve this problem, we consider a new set of candidate solutions of the form

$$\begin{aligned} \Theta = \{ \ \theta_{d,j} = (\tilde{\gamma}_d, \mathrm{e}^{\left(\frac{j}{10}-1\right)} \tilde{\lambda}_d) \ : \\ d = 1, \ldots, d_{\max}, j = 1, \ldots, j_{\max} \ \}. \end{aligned}$$

This family has cardinality $n_c = d_{\max} j_{\max}$. We take a large factor $\mathrm{e}^{\left(\frac{j}{10}-1\right)}$, to increase the probability of meeting the design specifications. Therefore, choosing a large enough value for $j_{\max}$ leads to a non-empty intersection of Θ with the set of parameters that meet the design specifications. In this example, we take $j_{\max} = 15$ and $d_{\max} = 10$. This yields to $n_c = 150$.

Using the finite family approach, we choose from Θ the design parameter that optimizes a given performance index. We draw from $\mathcal{W}$ a set V of N (i.i.d.) samples and select the pair that minimizes the empirical mean of the absolute value of the approximation error in the validation set V. That is, we consider the performance index

$$\frac{1}{N} \sum_{w \in V} \mathrm{e}^{\left(\frac{j}{10}-1\right)} \tilde{\lambda}_d^T |\varphi_d(w)|$$

subject to the constraints

$$|y(w) - \tilde{\gamma}_d^T \varphi_d(w)| \leq \mathrm{e}^{\left(\frac{j}{10}-1\right)} \tilde{\lambda}_d^T |\varphi_d(w)|, \ \forall w \in V.$$

We remark that the feasibility of this problem can be guaranteed in two ways. The first one is to choose $j_{\max}$ large enough. The second one is to allow m failures. As previously discussed, in this example we take $j_{\max} = 15$ and $m = 0$.

As the cardinality N of V has been chosen using Property 9, the probability of violation and the probability of failure of the best solution from Θ are bounded by η and δ respectively.

The obtained solution corresponds to $d = 5$ and $j = 13$. The corresponding value for the performance index is 0.8121. Figure 1 shows the approximation for the set V and the obtained probabilistic upper and lower bounds for the random function.

Finally, for illustrative purposes, we check with a validation set of sample size $N_v = 10N$, obtaining a number of 60 failures. The experimental violation probability turned out to be $\eta_{\exp} = 0.0146$, while the specification was $\eta = 0.05$.

9.2 Convex scenario approach

In this case we take advantage of the result of Subsection 9.1 and take $d = 5$ as the order of the approximation polynomial. Following the scenario approach we draw a set $\mathcal{W}_k$ of N samples (i.i.d) from $\mathcal{W}$ and solve the convex optimization problem

$$\begin{aligned} \min_{\gamma_d, \lambda_d} \quad & \lambda_d^T E\{|\varphi_d(t)|\} \\ s.t. \quad & |y(w) - \gamma_d^T \varphi_d(w)| \leq \lambda_d^T |\varphi_d(w)|, \ \forall w \in \mathcal{W}_k. \end{aligned}$$

In order to guarantee the design specifications we use Property 10 to determine the value of N. Since the number of decision variables is $2(d+1)$, $\eta = 0.05$ and $\delta = 10^{-6}$, the resulting value for N is 845. We notice that the convex scenario approach does not apply directly to the minimization of the empirical mean. This is why one has to resort to the exact computation of the mean of the approximation error $\lambda_d^T E\{|\varphi_d(t)|\}$, see [14].

Figure (1) shows the initial data set generated using the procedure described above, plus the envelope that contains all the solution polynomials.

Again, for illustrative purposes, we check with a validation set of size $N_v = 10N$, obtaining zero failures. The experimental value $\eta_{\exp} = 0$ is obtained, while the specification was $\eta = 0.05$.

Using this strategy, 845 design data are required, bigger than the number of required samples to use the finite families approach. We obtained a performance index of 0.8748, slightly larger than that obtained by the finite families strategy. The advantage of the finite families approach is that, using a smaller number of samples, a similar performance is obtained. This allows us to determine the best order of the polynomial with the further advantage that the exact computation of the mean of the error is not required.

9.3 SPV algorithm

We again take advantage of the result of Subsection 9.1 and take $d = 5$ as the order of the approximation polynomial. Following the SPV algorithm approach, we begin

setting $\eta = 0.05$, confidence $\delta = 10^{-6}$, scalars $a = 0.75$, $\alpha = 2$ and iteration index $k = 1$. The initial $\mathcal{W}_k$ is a set of 400 samples drawn from $\mathcal{W}$ according to probability $\Pr_{\mathcal{W}}$.

(i) A candidate solution $\hat{\theta}_k$ to the problem

$$\begin{aligned} \min_{\gamma_d,\lambda_d} \quad & \lambda_d^T \sum |\varphi_d(t)| \\ s.t. \quad & |y(w) - \gamma_d^T \varphi_d(w)| \leq \frac{1}{1.2}\lambda_d^T |\varphi_d(w)|, \ \forall w \in \mathcal{W}_k \end{aligned}$$

is obtained.

(ii) Set $m_k = \lfloor a(k) \rfloor$ and

$$M_k = \left\lceil \frac{1}{\eta}\left(m_k + \ln \frac{\xi(\alpha)k^\alpha}{\delta} + \sqrt{2m_k \ln \frac{\xi(\alpha)k^\alpha}{\delta}} \right) \right\rceil.$$

(iii) Obtain validation set $\mathcal{V}_k = \{v^{(1)}, \ldots, v^{(M_k)}\}$ drawing M_k i.i.d. validation samples from $\mathcal{W}$ according to the probability $\Pr_{\mathcal{W}}$.

(iv) If $\sum_{\ell=1}^{M_k} g(\hat{\theta}_k, v^{(\ell)}) \leq m_k$, then $\hat{\theta}_k$ is a probabilistic solution. The failure function is $g(\hat{\theta}_k, v^{(\ell)})$

$$g(\hat{\theta}_k, v^{(\ell)}) := \begin{cases} 0 \text{ if } |y(w) - \gamma_d^T \varphi_d(w)| \leq \lambda_d^T |\varphi_d(w)| \\ 1 \qquad\qquad \text{otherwise.} \end{cases}$$

(v) Exit if the exit condition is satisfied.

(vi) $k = k + 1$. $\mathcal{W}_k = \mathcal{W}_k \bigcup V_k$. Goto (i).

Figure (1) shows the initial data set generated using the procedure described above, and the envelope that contains all the solution polynomials. Using this strategy, 686 data are required. We obtained a performance index of 0.8877, slightly larger than that obtained by the finite families strategy.

The number of failures obtained in the last step of the algorithm is $m = 5$, being the empirical probability of failure $\eta = \frac{m}{M_k} = 0 < 0.05$.

We remark that if we set $a = 0$ in the algorithm, there are no allowed failures and this coincides with the approach studied in [24]. In this case, the algorithm did not find a solution for $\eta = 0.05$ and $M_k < 30000$.This is consistent with the results on Section 7.

In Table 1 the results of the three approaches are compared for different values of η.

η	N_{finite}	N_{convex}	N_{SPV}
0.2	190	212	582
0.1	378	423	543
0.05	754	845	686
0.02	1884	2113	2230
0.01	2366	4225	4061

10 Conclusion

In this paper, we have derived sample complexity results for various analysis and design problems related to uncertain systems. In particular, we provided new results which guarantee that a binomial distribution expression is smaller than a pre-specified value. These results are subsequently exploited for the analysis of worst-case performance and constraint violation. With regard to design problems we considered the case of finite families and the special case when the design problem can be recast as a robust convex optimization problem.

We also presented a general class of randomized algorithms based on probabilistic validation. We provided a strategy to adjust the cardinality of the validation sets to guarantee that the obtained solutions meet the probabilistic specifications. The proposed strategy is compared with other schemes from the literature and it has been shown that a strict validation strategy in which the design parameter has to satisfy the constraints for all the elements of the validation set might not be appropriate in some situations. We proved that the proposed approach does not suffer from this limitation because it allows the use of non strict validation test. As it has been shown in this paper, this relaxed scheme allows us to reduce, in some cases dramatically, the number of iterations required by the sequential algorithm. Another advantage of the proposed approach is that it does not rely on the existence of a robust deterministic solution. Furthermore, the presented strategy is quite general and is not based on a convexity assumption.

Acknowledgements

This work was supported by the MCYT-Spain and the European Commission which funded this work under projects DPI2010-21589-C05-01, FP7-257462 and HYCON2 Network of Excellence.

References

[1] T. Alamo, R. Tempo, D. R. Ramirez and E. F. Camacho. A sequentially optimal randomized algorithm for robust LMI feasibility problems. *Proceedings of the European Control Conference*, Kos, Greece, 2007.

[2] T. Alamo, R. Tempo, and E.F. Camacho. Randomized strategies for probabilistic solutions of uncertain feasibility and optimization problems. *IEEE Transactions on Automatic Control*, 54(11): 2545–2559, 2009.

[3] T. Alamo, R. Tempo, and A. Luque. On the sample complexity of probabilistic analysis and design methods. *Perspectives in Mathematical System Theory, Control, and Signal Processing*. Springer-Verlag, London, pp. 39-50, 2010

[4] T. Alamo, R. Tempo and A. Luque. On the sample complexity of randomized approaches to the analysis and design under uncertainty. *Proceedings of the American Control Conference*, Baltimore, USA, 2010.

[5] T. Alamo, A. Luque, D. R. Ramirez and R. Tempo. Randomized control design through probabilistic validation. *Proceedings of the American Control Conference*, Montreal, Canada, 2012.

[6] E. Bai, H. Cho, R. Tempo and Y. Ye. Optimization with few violated constraints for linear bounded error parameter estimation. *IEEE Transactions on Automatic Control*, 47(7):1067–1077, 2002.

[7] V. Blondel and J. N. Tsitsiklis. A survey of computational complexity results in system and control. *Automatica*, 36:1249-1274, 2000.

[8] G. Calafiore and B. T. Polyak, Stochastic algorithms for exact and approximate feasibility of robust LMIs. *IEEE Transactions on Automatic Control*, 46(11): 1755-1759, 2001.

[9] G. Calafiore and M. C. Campi. Uncertain convex programs: Randomized solutions and confidence levels. *Mathematical Programming*, 102:25-46, 2005.

[10] G. Calafiore and M.C. Campi. The scenario approach to robust control design. *IEEE Transactions on Automatic Control*, 51(5):742–753, 2006.

[11] G. Calafiore and F. Dabbene. A probabilistic analytic center cutting plane method for feasibility of uncertain LMIs. *Automatica*, 43:2022-2033, 2007.

[12] G. Calafiore. Random convex programs. *SIAM Journal of Optimization*, 20:3427-3464, 2010.

[13] G. Calafiore, F. Dabbene and R. Tempo. Research on probabilistic methods for control system design. *Automatica*, 47:1279-1293, 2011.

[14] M.C. Campi and S. Garatti. The exact feasibility of randomized solutions of robust convex programs. *SIAM Journal on Optimization*, 19:1211-1230, 2008.

[15] M. C. Campi and S. Garatti. A sampling-and-discarding approach to chance-constrained optimization: feasibility and optimality. *Journal of Optimization Theory and Applications*, 148: 257-280, 2011.

[16] H. Chernoff. A measure of asymptotic efficiency for tests of a hypothesis based on the sum of observations. *Annals of Mathematical Statistics*, 23:493–507, 1952.

[17] F. Dabbene, P. S. Shcherbakov and B. T. Polyak. A randomized cutting plane method with probabilistic geometric convergence. *SIAM Journal of Optimization*, 20:3185-3207, 2010.

[18] Y. Fujisaki and Y. Kozawa. Probabilistic robust controller design: probable near minimax value and randomized algorithms. In G. Calafiore and F. Dabbene, editors, *Probabilistic and Randomized Methods for Design under Uncertainty*. Springer, London, 2006.

[19] H. Ishii, T. Basar and R. Tempo Randomized algorithms for synthesis of switching rules for multimodal systems. *IEEE Transactions on Automatic Control*, Vol. 50, pp. 754-767, 2005.

[20] S. Kanev, B. De Schutter and M. Verhaegen. An ellipsoid algorithm for probabilistic robust controller design. *System and Control Lettets*, 49:365-375, 2003.

[21] V. Koltchinskii, C.T. Abdallah, M. Ariola, P. Dorato, and D. Panchenko. Improved sample complexity estimates for statistical learning control of uncertain systems. *IEEE Transactions on Automatic Control*, 45(12):2383–2388, 2000.

[22] D. Liberzon and R. Tempo Common Lyapunov functions and gradient algorithms *IEEE Transactions on Automatic Control*, Vol. 49, pp. 990-994, 2004.

[23] J. Luedtke, and S. Ahmed. A sample approximation approach for optimization with probabilistic constraints. *SIAM Journal of Optimization*, 19(2):674–699, 2008.

[24] Y. Oishi. Polynomial-time algorithms for probabilistic solutions of parameter-dependent linear matrix inequalities. *Automatica*, 43:538-545, 2007.

[25] M. Okamoto. Some inequalities relating to the partial sum of binomial probabilities. *Annals of the Institute of Statistical Mathematics*, 10(1):29–35, 1959.

[26] B. T. Polyak and R. Tempo. Probabilistic robust design with linear quadratic regulators. *Systems and Control Letters*, 43:343-353, 2001.

[27] C. Scherer. LMI relaxations in robust control. *European Journal of Control*, 12:3-29, 2006.

[28] T. Wada and Y. Fujisaki. Probabilistic cutting plane technique based on Maximum Volume Ellipsoid Center *Proceedings of the 48th IEEE Conference Decision and Control and the 28th Chinese Control Conference*, pp. 1169-1174, Shanghai, China, (2009)

[29] T. Wada and Y. Fujisaki. Stochastic ellipsoid methods for robust control: Multiple Updates and Multiple Cuts, *Automatica*, Vol. 46, No. 8, pp. 1339-1345 (2010)

[30] R. Tempo, E.W. Bai, and F. Dabbene. Probabilistic robustness analysis: explicit bounds for the minimum number of samples. *Systems & Control Letters*, 30:237–242, 1997.

[31] R. Tempo, G. Calafiore, and F. Dabbene. *Randomized Algorithms for Analysis and Control of Uncertain Systems, with Applications*. Second Edition, Springer-Verlag, London, 2013.

[32] V.N. Vapnik. *Statistical Learning Theory*. John Wiley & Sons, New York, 1998.

[33] M. Vidyasagar. *A Theory of Learning and Generalization: with Applications to Neural Networks and Control Systems*. Springer, London, 1997.

[34] M. Vidyasagar. Randomized algorithms for robust controller synthesis using statistical learning theory. *Automatica*, 37:1515-1528, 2001.

[35] M. Vidyasagar and V. D. Blondel. Probabilistic solutions to some NP-hard matrix problems. *Automatica* , 37:1397-1405, 2001.

11 Appendix

Proof of Corollary 1

We first consider the case $m = 0$. Then, we obtain $B(N, \eta, 0) = (1-\eta)^N = \mathrm{e}^{N \ln(1-\eta)} \le \mathrm{e}^{-\eta N}$. Therefore,

it follows from $\eta N \geq \ln \frac{1}{\delta}$ that $\mathrm{e}^{-\eta N} \leq \mathrm{e}^{\ln \delta} = \delta$. This proves the result for $m = 0$.

Consider now the case $m > 0$. We first prove that

$$h(r) := \sqrt{2(r-1)} - \ln\left(r + \sqrt{2(r-1)}\right) \geq 0, \quad \forall r \geq 1. \tag{15}$$

Since $h(1) = 0$, the inequality $h(r) \geq 0$ holds if the derivative of $h(r)$ is strictly positive for every $r > 1$.

$$\frac{\mathrm{d}}{\mathrm{d}r} h(r) =$$

$$\frac{1}{\sqrt{2(r-1)}} - \frac{1}{r + \sqrt{2(r-1)}}\left(1 + \frac{1}{\sqrt{2(r-1)}}\right) =$$

$$\left(\frac{1}{\sqrt{2(r-1)}}\right)\left(1 - \frac{1 + \sqrt{2(r-1)}}{r + \sqrt{2(r-1)}}\right) =$$

$$\left(\frac{1}{\sqrt{2(r-1)}}\right)\left(\frac{r-1}{r + \sqrt{2(r-1)}}\right) \geq 0, \quad \forall r > 1.$$

This proves the inequality $h(r) \geq 0$, for all $r \geq 1$. Denote now $\hat{a} = r + \sqrt{2(r-1)}$, with $r = 1 + \frac{1}{m} \ln \frac{1}{\delta}$. Clearly $\hat{a} > 1$. Therefore, from a direct application of Lemma 2, we conclude that it suffices to choose N such that

$$N\eta \geq \frac{\hat{a}}{\hat{a} - 1}\left(\ln \frac{1}{\delta} + m \ln \hat{a}\right) =$$

$$\frac{r + \sqrt{2(r-1)}}{r - 1 + \sqrt{2(r-1)}}\left(r - 1 + \ln\left(r + \sqrt{2(r-1)}\right)\right) m.$$

Since $h(r) \geq 0$ we conclude that

$$\frac{r - 1 + \ln\left(r + \sqrt{2(r-1)}\right)}{r - 1 + \sqrt{2(r-1)}} \leq 1.$$

From this inequality, we finally conclude that inequality $B(N, \eta, m) \leq \delta$ holds if

$$N\eta \geq (r + \sqrt{2(r-1)})m = m + \ln \frac{1}{\delta} + \sqrt{2m \ln \frac{1}{\delta}}.$$

$\square$

Property 4 *Suppose that N is a positive integer and that s is a strictly positive scalar. Then,*

$$\sum_{k=1}^{N} \frac{1}{k^s} \leq \Phi(s, \lceil \log_2 N \rceil)$$

where, given $s \geq 0$ and the integer $t \geq 0$,

$$\Phi(s,t) := \begin{cases} \dfrac{1 - 2^{(1-s)(t+1)}}{1 - 2^{1-s}} & \text{if } s \neq 1 \\ \\ t + 1 & \text{otherwise.} \end{cases}$$

Proof: Given $N > 0$ and $s > 0$, define $t := \lceil \log_2(N) \rceil$ and $S(t) := \sum\limits_{k=1}^{2^t} \frac{1}{k^s}$. With these definition we have

$$\sum_{k=1}^{N} \frac{1}{k^s} \leq \sum_{k=1}^{2^t} \frac{1}{k^s} = S(t).$$

In what follows we show that $S(t) \leq 1 + 2^{1-s} S(t-1)$ for every integer t greater than 0. Since $S(0) = 1$ and $S(1) = 1 + 2^{-s}$, the inequality is clearly satisfied for $t = 1$. We now prove the inequality for t greater than 1

$$\begin{aligned}
S(t) &= \sum_{k=1}^{2^t} \frac{1}{k^s} = \sum_{k=1}^{2^{t-1}} \left[\frac{1}{(2k)^s} + \frac{1}{(2k-1)^s}\right] \\
&= 2^{-s} \sum_{k=1}^{2^{t-1}} \frac{1}{k^s} + \sum_{k=1}^{2^{t-1}} \frac{1}{(2k-1)^s} \\
&= 2^{-s} S(t-1) + 1 + \sum_{k=2}^{2^{t-1}} \frac{1}{(2k-1)^s} \\
&\leq 2^{-s} S(t-1) + 1 + \sum_{k=2}^{2^{t-1}} \frac{1}{(2k-2)^s} \\
&= 2^{-s} S(t-1) + 1 + 2^{-s} \sum_{k=2}^{2^{t-1}} \frac{1}{(k-1)^s} \\
&= 2^{-s} S(t-1) + 1 + 2^{-s} \sum_{k=1}^{2^{t-1}-1} \frac{1}{k^s} \\
&\leq 2^{-s} S(t-1) + 1 + 2^{-s} \sum_{k=1}^{2^{t-1}} \frac{1}{k^s} \\
&= 2^{-s} S(t-1) + 1 + 2^{-s} S(t-1) \\
&= 1 + 2^{1-s} S(t-1).
\end{aligned}$$

We have therefore proved the inequality $S(t) \leq 1 + 2^{1-s} S(t-1)$ for every integer t greater than 0. Using this inequality in a recursive way with $S(0) = 1$ we obtain

$$S(t) \leq \sum_{k=0}^{t} 2^{(1-s)k} = \Phi(s,t).$$

This proves the result. $\square$